A

PATENTS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Li How Chen and Harry Meade

Serial No.: To be assigned

Filing Date: HEREWITH

Title: NOVEL MODIFIED NUCLEIC ACID SEQUENCES AND METHODS FOR INCREASING mRNA LEVELS AND PROTEIN EXPRESSION IN CELL SYSTEMS

Docket Number: 107.637.121-B

BOX PATENT APPLICATION
Assistant Commissioner for Patents
Washington, DC 20231

CERTIFICATION UNDER 37 CFR 1.10

I hear by certify that the attached papers are being deposited with the United States Postal Service as "Express Mail Post Office to Addressee" Service under 37 CFR 1.10 on October 20, 1998 and is addressed to: ***BOX PATENT APPLICATION****, Assistant Commissioner for Patents, Washington, D.C. 20231*

"Express Mail Label No."

Angela J. Mignanelli

TRANSMITTAL LETTER

Dear Sir:

Enclosed herewith for filing in the United States Patent and Trademark Office for the above-referenced application are the following documents:

1. New United States Patent Application filed under 37 C.F.R. § 1.53(b) entitled:

NOVEL MODIFIED NUCLEIC ACID SEQUENCES AND METHODS FOR INCREASING mRNA LEVELS AND PROTEIN EXPRESSION IN CELL SYSTEMS

and naming as inventors:

CHEN, Li How

MEADE, Henry

the application contains 42 pages comprising:

20 pages of the specification;

5 pages of the claims (26 claims which 10 are independent and 4 are multiple-dependent claims);

1 page of the abstract; and

16 sheets of informal drawings comprising Figures 1 to 13.

2. Small Entity Statement (unexecuted).

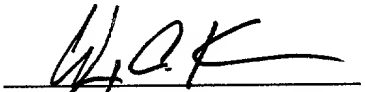
3. Return postcard.

This application claims priority to U.S.S.N. 60/095,649 filed on the 15 May 1998 and U.S.S.N. 60/062,592 filed 20 October 1997.

Please charge our deposit account, 08-0219, the required filing fee of $395.00 pursuant to 37 C.F.R. §1.16(a) , the required excess independent claims fee of $287.00 pursuant to 37 C.F.R. §1.16(b), and the required multiple dependent claim fee of $540.00 pursuant to 37 C.F.R. §1.16(d). The total amount of fees to be charged to our deposit account is $1222.00 For this purpose we enclose a duplicate copy of this document. The fees charged reflect the small entity status of this application.

Respectfully submitted,

HALE AND DORR LLP



Wayne A. Keown, Ph.D.
Registration No. 33,923
Attorney for Applicants

60 State Street
Boston, MA 02109
Tel: (617) 526-6000
Fax: (617) 526-5000

Date: October 20, 1998

NOVEL MODIFIED NUCLEIC ACID SEQUENCES AND METHODS FOR INCREASING mRNA LEVELS AND PROTEIN EXPRESSION IN CELL SYSTEMS

BACKGROUND OF THE INVENTION

Field of the invention

The invention relates to heterologous gene expression. More particularly, the invention relates to the expression of microbial or parasitic organism genes in higher eukaryote cell systems.

Summary of the related art

Recombinant production of certain heterologous gene products is often difficult in *in vitro* cell culture systems or *in vivo* recombinant production systems. For example, many researchers have found it difficult to express proteins derived from bacteria, parasites and virus in cell culture systems different from the cell from which the protein was originally derived, and particularly in mammalian cell culture systems. One example of a therapeutically important protein which has been difficult to produce by mammalian cells is the malaria merozoite surface protein (MSP-1).

Malaria is a serious heath problem in tropical countries. Resistance to existing drugs is fast developing and a vaccine is urgently needed. Of the number of antigens that get expressed during the life cycle of *P. falciparum*, MSP-1 is the most extensively studied and promises to be the most successful candidate for vaccination. Individuals exposed to *P. falciparum* develop antibodies against MSP-1, and studies have shown that there is a correlation between a naturally acquired immune response to MSP-1 and reduced malaria morbidity. In a number of studies, immunization with purified native MSP-1 or recombinant fragments of the protein has induced at least partial protection from the parasite (Diggs et al, (1993) *Parasitol. Today* 9:300-302). Thus MSP-1 is an important target for the

development of a vaccine against *P. falciparum.*

MSP-1 is a 190-220 kDA glycoprotein. The C-terminal region has been the focus of recombinant production for use as a vaccine. However, a major problem in developing MSP-1 as a vaccine is the difficulty in obtaining recombinant proteins in bacterial or yeast expression systems that are equivalent in immunological potency to the affinity purified native protein (Chang et al., (1992) *J. Immunol.* **148**:548-555.) and in large enough quantities to make vaccine production feasible.

Improved procedures for enhancing expression of sufficient quantities of proteins derived from parasite, bacterial and viral organisms which have previously been difficult to produce recombinantly would be advantageous. In particular, a recombinant system capable of expressing MSP-1 in sufficient quantities would be particularly advantageous.

BRIEF SUMMARY OF THE INVENTION

The present invention provides improved recombinant DNA compositions and procedures for increasing the mRNA levels and protein expression of proteins derived from heterologous cells, preferably those of lower organisms such as bacteria, virus, and parasite, which have previously been difficult to express in cell culture systems, mammalian cell culture systems, or in transgenic mammals. The preferred protein candidates for expression in an expression system in accordance with the invention are those proteins having DNA coding sequences comprising high overall AT content or AT rich regions, and/or mRNA instability motifs and/or rare codons relative to the recombinant expression systems.

In a first aspect, the invention features a modified known nucleic acid, preferably a gene from a bacterium, virus or parasite, capable of being expressed in a system, wherein the modification comprises a reduced AT content, relative to the unmodified sequence, and optionally further comprises elimination of at least one or all mRNA instability motifs present in the natural gene. In certain preferred embodiments the modification further comprises replacement of one or more codons of the natural gene with preferred codons of the cell system.

In a second aspect, the invention provides a process for preparing a modified nucleic acid of the invention comprising the steps of lowering the overall AT content of the natural gene encoding the protein, and/or eliminating at least one or all mRNA instability motifs and/or replacing one or more codons with a preferred codon of the cell system of choice, all by replacing one or more codons in the natural gene with codons recognizable to, and preferably with codons preferred by the cell system of choice and which code for the same amino acids as the replaced codon. This aspect of the invention further includes modified nucleic acids prepared according to the process of the invention.

In a third aspect, the invention also provides vectors comprising nucleic acids of the invention and promoters active in the cell line or organism of choice, and

host cells transformed with nucleic acids of the invention.

In a fourth aspect, he invention provides transgenic expression vectors for the production of transgenic lactating animals comprising nucleic acids of the invention as well as transgenic non-human lactating animals whose germlines comprise a nucleic acid of the invention.

In a fifth aspect, he invention provides a transgenic expression vector for production of a transgenic lactating animal species comprising a nucleic acid of the invention, a promoter operatively coupled to the nucleic acid which directs mammary gland expression of the protein encoded by the nucleic acid into the milk of the transgenic animal.

In a sixth aspect, the invention provides a DNA vaccine comprising a modified nucleic acid according to the invention. A preferred embodiment of this aspect of the invention comprises a fragment of a modified MSP-1 gene according to the invention.

DESCRIPTION OF THE DRAWINGS

Fig. 1 depicts the cDNA sequence of MSP-1_{42} modified in accordance with the invention [SEQ ID NO 1] in which 306 nucleotide positions have been replaced to lower AT content and eliminate mRNA instability motifs while maintaining the same protein amino acid sequence of MSP-1_{42}. The large letters indicate nucleotide substitutions.

Fig. 2 depicts the nucleotide sequence coding sequence of the "wild type" or native MSP- 1_{42} [SEQ ID NO 2].

Fig 3a is a codon usage table for wild type MSP-1_{42} (designated "MSP wt" in the table) and the new modified MSP-1_{42} gene (designated "edited MSP" in the table) and several milk protein genes (casein genes derived from goats and mouse). The numbers in each column indicate the actual number of times a specific codon appears in each of the listed genes. The new MSP-1_{42} synthetic gene was derived from the mammary specific codon usage by first choosing GC rich codons for a given amino acid combined with selecting the amino acids used most frequently in the milk proteins.

Fig 3b is a codon usage table comparing the number of times each codon appears in both the wild type MSP-1_{42} (designated "MSP wt" in the table) and the new modified MSP-1_{42}.gene (designated "edited MSP" in the table) as is also shown in the table in Fig. 3a. The table in Fig. 3b, also compares the frequency in which each codon appears in the wild type MSP-1_{42} and the new modified MSP-1_{42} gene, to the frequency of appearance of each codon in both *E.coli* genes and human genes. Thus, if the expression system were *E.coli* cells, this table may be used to determine what codons are recognized by, or preferred by *E.coli*.

Fig. 4a-c depict MSP-1_{42} constructs GTC 479, GTC 564, and GTC 627, respectively as

are described in the examples.

Fig. 5 panel A is a Northern analysis wherein construct GTC627 comprises the new MSP-1_{42} gene modified in accordance with the invention, GTC479 is the construct comprising the native MSP-1_{42} gene, and construct GTC469 is a negative control DNA

Fig 5 panel B is a Western analysis wherein the eluted fractions after affinity purifications. Numbers are collected fractions. The results show that fractions from GTC679 the modified MSP-1_{42} synthetic gene construct reacted with polyclonal antibodies to MSP-1 and the negative control GTC479 did not.

Fig 6 depicts the nucleic acid sequences of OT1 [SEQ ID NO 3], OT2 [SEQ ID NO 4], MSP-8 [SEQ ID NO 5] MSP-2 [SEQ ID NO 6] and MSP1 [SEQ ID NO 7] described in the Examples.

Fig 7 is a schematic representation of plasmid BC574.

Fig 8 is a schematic representation of BC620.

Fig 9 is a schematic representation of BC670.

Fig 10 is a representation of a Western blot of MSP in transgenic milk.

Fig 11 is a schematic representation of the nucleotide sequence of MSP42-2 [SEQ ID NO 8].

Fig 12 is a schematic representation of the BC-718.

Fig 13 is a representation of a Western blot of BC-718 expression in transgenic milk.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The patent and scientific literature referred to herein establishes the knowledge that is available to those with skill in the art. The issued US patents, allowed applications, published foreign applications, and references cited herein are hereby incorporated by reference. Any conflicts between these references and the present disclosure shall be resolved in favor of the present disclosure.

The invention provides modified recombinant nucleic acid sequences (preferably DNA) and methods for increasing the mRNA levels and protein expression of proteins which are known to be, or are likely to be, difficult to express in cell culture systems, mammalian cell culture systems, or in transgenic animals. The preferred "difficult" protein candidates for expression using the recombinant techniques of the invention are those proteins derived from heterologous cells preferably those of lower organisms such as parasites, bacteria, and virus, having DNA coding sequences comprising high overall AT content or AT rich regions and/or mRNA instability motifs and/or rare codons relative to the recombinant expression system to be used.

In a first aspect, the invention features a modified known nucleic acid, preferably a gene from a bacterium, virus or parasite, capable of being expressed in a cell system, wherein the modification comprises a reduced AT content, relative to the unmodified sequence, and optionally further comprises elimination of at least one or all mRNA instability motifs present in the natural gene. A "cell system" includes cell culture systems, tissue culture systems, organ culture systems and tissues of living animals. In certain preferred embodiments the modification further comprises replacement of one or more codons of the natural gene with preferred codons of the cell system. Each of these features are achieved by replacing one or more codons of the natural gene with codons recognizable to, and preferably preferred by the cell system that encode the same amino acid as the codon which

was replaced in the natural gene. In accordance with the invention, such "silent" nucleotide and codon substitutions should be sufficient to achieve the goal lowering AT content and/or of eliminating mRNA instability motifs, and/or reducing the number of rare codons, while maintaining, and preferably improving the ability of the cell system to produce mRNA and express the desired protein.

Also included in the invention are those sequences which are specifically homologous to the modified nucleic acids of the invention under suitable stringent conditions, specifically excluding the known nucleic acids from which the modified nucleic acids are derived. A sequence is "specifically homologous" to another sequence if it is sufficiently homologous to specifically hybridize to the exact complement of the sequence. A sequence "specifically hybridizes" to another sequence if it hybridizes to form Watson-Crick or Hoogsteen base pairs either in the body, or under conditions which approximate physiological conditions with respect to ionic strength, *e.g.*, 140 mM NaCl, 5 mM $MgCl_2$. Preferably, such specific hybridization is maintained under stringent conditions, *e.g.*, 0.2X SSC at 68°C.

In preferred embodiments, the nucleic acid of the invention is capable of expressing the protein in mammalian cell culture, or in a transgenic animal at a level which is at least 25%, and preferably 50% and even more preferably at least 100% or more of that expressed by the natural gene in an in vitro cell culture system or in a transgenic animal under identical conditions (i.e. the same cell type, same culture conditions, same expression vector).

As used herein, the term "expression" is meant mRNA transcription resulting in protein expression. Expression may be measured by a number of techniques known in the art including using an antibody specific for the protein of interest. By "natural gene" or "native gene" is meant the gene sequence, or fragments thereof (including naturally occurring allelic variations), which encode the wild type form of the protein and from which the modified nucleic acid is derived. A "preferred codon "means a codon which is used more prevalently by the cell system of choice. Not all codon changes described herein are changes to a

preferred codon, so long as the codon replacement is a codon which is at least recognized by the cell system. The term "reduced AT content" as used herein means having a lower overall percentage of nucleotides having A (adenine) or T (thymine) bases relative to the natural gene due to replacement of the A or T containing nucleotide positions or A and/or T containing codons with nucleotides or codons recognized by the cell system of choice and which do not change the amino acid sequence of the target protein. "Heterologous" is used herein to denote genetic material originating from a different species than that into which it has been introduced, or a protein produced from such genetic material.

Particularly preferred cell systems of the invention include mammalian cell culture systems such as COS cells and CHO cells, as well as transgenic animals, particularly the mammary tissue of transgenic animals. However, the invention also contemplates bacteria, yeast, E. coli, and viral expression systems such as baculovirus and even plant systems.

In a second aspect, the invention provides a process for preparing a modified nucleic acid of the invention comprising the steps of lowering the overall AT content of the natural gene encoding the protein, and/or eliminating at least one or all mRNA instability motifs and/or replacing one or more codons with a preferred codon of the cell system of choice, all by replacing one or more codons in the natural gene with codons recognizable to, and peferably with codons preferred by the cell system of choice and which code for the same amino acids as the replaced codon. Standard reference works describing the general principals of recombinant DNA technology include Watson, J.D. et al, *Molecular Biology of the Gene*, Volumes I and II the Benjamin/Cummings Publishing Company, Inc. publisher, Menlo Park, CA (1987) Darnell, J.E. et al., *Molecular Cell Biology*, Scientific American Books, Inc., Publisher, New York, NY (1986); Old, R.W., et al., *Principles of Gene Manipulation: An Introduction to Genetic Engineering*, 2d edition, University of California Press, publisher, Berkeley CA (1981); Maniatis, T., et al., *Molecular Cloning: A Laboratory*

Manual, 2nd ed. Cold Spring Harbor Laboratory, publisher, Cold Spring Harbor, NY (1989) and *Current Protocols in Molecular Biology*, Ausubel et al., Wiley Press, New York, NY (1992). This aspect of the invention further includes modified nucleic acids prepared according to the process of the invention.

Without being limited to any theory, previous research has indicated that a conserved AU sequence (AUUUA) from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation (Shaw, G. and Kamen, R. *Cell* **46**:659-667). The focus in the past has been on the presence of these instability motifs in the untranslated region of a gene. The instant invention is the first to recognize an advantage to eliminating the instability sequences in the coding region of a gene.

In a third aspect, the invention also provides vectors comprising nucleic acids of the invention and promoters active in the cell line or organism of choice, and host cells transformed with nucleic acids of the invention. Preferred vectors include an origin of replication and are thus replicatable in one or more cell type. Certain preferred vectors are expression vectors, and further comprise at least a promoter and passive terminator, thereby allowing transcription of the recombinant expression element in a bacterial, fungal, plant, insect or mammalian cell.

In a fourth aspect, he invention provides transgenic expression vectors for the production of transgenic lactating animals comprising nucleic acids of the invention as well as transgenic non-human lactating animals whose germlines comprise a nucleic acid of the invention. Such transgenic expression vectors comprise a promoter capable of being expressed as part of the genome of the host transgenic animal. General principals for producing transgenic animals are known in the art. See for example Hogan et al., *Manipulating the Mouse Embryo: A Laboratory Manual*, Cold Spring Harbor Laboratory, (1986); Simons et al, *Bio/Technology* **6**:179-183, (1988); Wall et al., *Biol. Reprod.* **32**:645-651, (1985); Buhler et al., *Bio/Technology*, **8**:140-143 (1990); Ebert et al., *Bio/Technology* **9**:835-838 (1991);

Krimenfort et al., *Bio/Technology* **9**:844-847 (1991); Wall et al., *J.Cell. Biochem.* **49**:113-120 (1992). Techniques for introducing foreign DNA sequences into mammals and their germ cells were originally developed in the mouse. See e.g., Gordon et al., *Proc. Natl. Acad. Sci. USA* **77**:7380-7384, (1980); Gordon and Ruddle, *Science* **214**: 1244-1246 (1981); Palmiter and Brinster, *Cell* **41**: 343-345, 1985; Brinster et al., *Proc Natl. Acad Sci., USA* **82**:4438-4442 (1985) and Hogan et al. (*ibid.*). These techniques were subsequently adapted for use with larger animals including cows and goats. Up until very recently, the most widely used procedure for the generation of transgenic mice or livestock, several hundred linear molecules of the DNA of interest in the form of a transgenic expression construct are injected into one of the pro- nuclei of a fertilized egg. Injection of DNA into the cytoplasm of a zygote is also widely used. Most recently cloning of an entire transgenic cell line capable of injection into an unfertilized egg has been achieved (KHS Campbell et al., *Nature* **380** 64-66, (1996)).

In a fifth aspect, he invention provides a transgenic expression vector for production of a transgenic lactating animal species comprising a nucleic acid of the invention, a promoter operatively coupled to the nucleic acid which directs mammary gland expression of the protein encoded by the nucleic acid into the milk of the transgenic animal. The mammary gland expression system has the advantages of high expression levels, low cost, correct processing and accessibility. Known proteins, such as bovine and human alpha- lactalbumin have been produced in lactating transgenic animals by several researchers. (Wright et al, *Bio/Technology* **9**:830-834 (1991); Vilotte et al, *Eur. J. Biochem.*,**186**:43-48 (1989); Hochi et al., *Mol Reprod. And Devel.* **33**:160-164 (1992); Soulier et al., *FEBS Letters* **297**(1,2):13-18 (1992)) and the system has been shown to produce high levels of protein.

Preferred promoters are active in the mammary tissue. Particularly useful are promoters that are specifically active in genes encoding milk specific proteins such

as genes found in mammary tissue, i.e. are more active in mammary tissue than in other tissues under physiological conditions where milk is synthesized. Most preferred are promoters that are both specific to and efficient in mammary tissue. Among such promoters, the casein, lactalbumin and lactalglobulin promoters are preferred, including, but not limited to the alpha, beta and gamma casein promoters and the alpha lactalbumin and beta-lactalglobulin promoters. Preferred among the promoters are those from rodent, goats and cows. Other promoters include those that regulate a whey acidic protein (WAP) gene.

In a preferred embodiment of the invention, a modified nucleic acid encoding MSP-1 or fragments thereof capable of expression in a cell culture system, mammalian cell culture system or in the milk of a transgenic animal is provided. Nucleic acid sequences encoding the natural MSP-1 gene are modified in accordance with the invention. First the overall AT content is reduced by replacing codons of the natural gene with codons recognizable to, and preferably with codons preferred by the cell system of choice, that encode the same amino acid but are sufficient to lower the AT content of the modified nucleic acid as compared to the native MSP-1 gene or gene fragment. Second, mRNA instability motifs (AUUUA, Shaw and Kamen, *supra*) in the native gene or gene fragment are eliminated from the coding sequence of the gene by replacing codons of the natural gene with codons recognizable to, and preferably prefrred by the cell system of choice that encode the same amino acid but are sufficient to eliminate the mRNA instability motif. Optionally, any other codon of the native gene may be replaced with a preferred codon of the expression system of choice as described.

In a sixth aspect, the invention provides a DNA vaccine comprising a modified nucleic acid according to the invention. In certain preferred embodiments, the DNA vaccine conprises a vector according to the invention, The DNA vaccine according to the invention may be in the form of a "naked" or purified modified nucleic acid according to the invention, which may or may not be

operatively associated with a promoter. A nucleic acid is operatively associated with a promoter if it is associated with the promoter in a manner which allows the nucleic acid sequence to be expressed. Such DNA vaccines may be delivered without encapsulation, or they may be delivered as part of a liposome, or as part of a viral genome. Generally, such vaccines are delivered in an amount sufficient to allow expression of the nucleic acid and elicit an antibody response in an animal, including a human, which receives the DNA vaccine. Subsequent deliveries, at least one week after the first delivery, may be used to enhance the antibody response. Preferred delivery routes include introduction via mucosal membranes, as well as parenteral administration.

A preferred embodiment of this aspect of the invention comprises a fragment of a modified MSP-1 gene according to the invention. Such fragment preferably includes from about 5% to about 100% of the overall gene sequence and comprises one or more modification according to the invention.

Examples of codon usage from E.coli and human are shown in Fig. 3b. Fig. 3b shows the frequency of codon usage for the MSP-1 native gene as well as the modified MSP-1 gene of the invention and also compares the frequency of codon usage to that of E. coli and human genes. Codon usage frequency tables are readily available and known to those skilled in the art for a number of other expression systems such as yeast, baculovirus and the mammalian, systems.

The following examples illustrate certain preferred modes of making and practicing the present invention, but are not meant to limit the scope of the invention since alternative methods may be utilized to obtain similar results.

Examples

Creation of novel modified MSP-1_{42} gene

In one embodiment, a novel modified nucleic acid encoding the C-terminal fragment of MSP-1 is provided. The novel, modified nucleic acid of the invention

encoding a 42 kD C-terminal part of MSP-1 (MSP-1_{42}) capable of expression in mammalian cells of the invention is shown in Fig. 1. The natural MSP-1_{42} gene (Fig 2) was not capable of being expressed in mammalian cell culture or in transgenic mice Analysis of the natural MSP- 1_{42} gene suggested several characteristics that distinguish it from mammalian genes. First, it has a very high overall AT content of 76%. Second, the mRNA instability motif, AUUUA, occurred 10 times in this 1100 bp DNA segment (Fig 2). To address these differences a new MSP-1_{42} gene was designed. Silent nucleotide substitution was introduced into the native MSP-1_{42} gene at 306 positions to reduce the overall AT content to 49.7%. Each of the 10 AUUUA mRNA instability motifs in the natural gene were eliminated by changes in codon usage as well. To change the codon usage, a mammary tissue specific codon usage table, Fig. 3a, was created by using several mouse and goat mammary specific proteins. The table was used to guide the choice of codon usage for the modified MSP-1_{42} gene as described above. For example as shown in the Table in Fig. 3a, in the natural gene, 65% (25/38) of the Leu was encoded by TTA, a rare codon in the mammary gland. In the modified MSP-1_{42} gene, 100% of the Leu was encoded by CTG, a preferred codon for Leu in the mammary gland.

An expression vector was created using the modified MSP-1_{42} gene by fusing the first 26 amino acids of goat beta-casein to the N-terminal of the modified MSP-1_{42} gene and a SalI-Xho I fragment which carries the fusion gene was subcloned into the XhoI site of the expression vector pCDNA3. A His6 tag was fused to the 3′ end of the MSP-1_{42} gene to allow the gene product to be affinity purified. This resulted in plasmid GTC627 (Fig.4c).

To compare the natural MSP-1_{42} gene construct to the modified MSP-1_{42} nucleic acid of the invention, an expression vector was also created for the natural MSP-1_{42} gene and the gene was added to mammalian cell culture and injected into mice to

form transgenic mice as follows:

Construction of the native MSP-1_{42} *Expression Vector*

To secrete the truncated the merozoite surface protein-1 (MSP-1) of Plasmodium falciparum, the wild type gene encoding the 42KD C-terminal part of MSP-1 (MSP-1_{42}) was fused to either the DNA sequence that encodes the first 15 or the first 25 amino acids of the goat beta-casein. This is achieved by first PCR amplify the MSP-1 plasmid (received from Dr. David Kaslow, NIH) with primers MSP1 and MSP2 (Fig. 6), then cloned the PCR product into the TA vector (Invitrogen). The Bg1II-XhoI fragments of the PCR product was ligated with oligos OT1 and OT2 (Fig. 6) into the expression vector pCDNA3. This yielded plasmid GTC564 (Fig.4b), which encodes the 15 amino acid beta- casein signal peptide and the first 11 amino acids of the mature goat beta-casein followed by the native MSP-1_{42} gene. Oligos MSP-8 and MSP-2 (Fig. 6) were used to amplify MSP-1 plasmid by PCR, the product was then cloned into TA vector. The XhoI fragment was exercised and cloned into the XhoI site of the expression vector pCDNA3 to yield plasmid GTC479 (Fig.4a), which encoded 15 amino acid goat beta-casein signal peptide fused to the wild-type MSP-1_{42} gene. A His6 tag was added to the 3′ end of MSP-1_{42} gene in GTC 564 and GTC 479.

Native MSP-1_{42} Gene Is Not Expressed In COS-7 Cells

Expression of the native MSP gene in cultured COS-7 cells was assayed by transient transfection assays. GTC479 and GTC564 plasmids DNA were introduced into COS-7 cells by lipofectamine (Gibco-BRL) according to manufacturer's protocols. Total cellular RNA was isolated from the COS cells two days post-transfection. The newly synthesized proteins were metabolically labeled for 10 hours by adding ^{35}S methionine added to the culture media two days-post transfection.

To determine the MSP mRNA expression in the COS cells, a Northern blot was probed with a ^{32}P labeled DNA fragment from GTC479. No MSP RNA was detected in GTC479 or GTC564 transfectants (data not shown). Prolonged exposure revealed residual levels of degraded MSP mRNA. The ^{35}S labeled culture supernatants and the lysates were immunoprecipitated with a polyclonal antibody raised against MSP. Immunoprecipitation experiments showed that no expression from either the lysates or the supernatants of the GTC479 or GTC564 transfected cells (data not shown). These results showed that the native MSP-1 gene was not expressed in COS cells.

Native MSP-1_{42} Gene is Not Expressed in the Mammary Gland of Transgenic Mice

The SalI-XhoI fragment of GTC479, which encoded the 15 amino acids of goat beta-casein signal peptide, the first 11 amino acids of goat beta-casein, and the native MSP-1_{42} gene, was cloned into the XhoI site of the beta-casein expressed in vector BC350. This yielded plasmid BC574 (Fig.7). A SalI-NotI fragment of BC574 was injected into the mouse embryo to generate transgenic mice. Fifteen lines of transgenic mice were established. Milk from the female founder mice was collected and subjected to Western analysis with polycolonal antibodies against MSP. None of the seven mice analyzed were found to express MSP-1_{42} protein in their milk. To further determine if the mRNA of MSP-1_{42} was expressed in the mammary gland, total RNA was extracted from day 11 lactating transgenic mice and analyzed by Northern blotting. No MSP-1_{42} mRNA was detected by any of the BC 574 lines analyzed. Therefore, the MSP-1_{42} transgene was not expressed in the mammary gland of transgenic mice. Taken together, these experiments suggest that native parasitic MSP-1_{42} gene could not be expressed in mammalian cells, and the block is as the level of mRNA abundance.

Expression of MSP in the Mammalian Cells

Transient transfection experiments were performed to evaluate the expression of the modified MSP-1_{42} gene of the invention in COS cells. GTC627 and GTC479 DNA were introduced into the COS-7 cells. Total RNA was isolated 48 hours post-transfection for Northern analysis. The immobilized RNA was probed with ^{32}P labeled SalI-XhoI fragment of GTC627. A dramatic difference was observed between GTC479 and GTC627. While no MSP-1_{42} mRNA was detected in the GTC479 transfected cells as shown previously, abundant MSP-1_{42} mRNA was expressed by GTC627 (Fig. 5, Panel A). GTC 469 was used as a negative control and comprises the insert of GTC564 cloned into cloning vector PU19, a commercially available cloning vector. A metabolic labeling experiment with ^{35}S methionine followed by immunoprecipitation with polyclonal antibody (provided by D. Kaslow NIAID, NIH) against MSP showed that MSP-1_{42} protein was synthesized by the transfected COS cells (Fig.5, Panel B). Furthermore, MSP-1_{42} was detected in the transfected COS supernatant, indicating the MSP-1_{42} protein was also secreted. Additionally, using Ni-NTA column, MSP-1_{42} was affinity purified from the GTC627 transfected COS supernatant.

These results demonstrated that the modification of the parasitic MSP-1_{42} gene lead to the expression of MSP mRNA in the COS cells. Consequently, the MSP-1_{42} product was synthesized and secreted by mammalian cells.

Polyclonal antibodies used in this experiment may also be prepared by means well known in the art (*Antibodies: A Laboratory Manual*, Ed Harlow and David Lane, eds. Cold Spring Harbor Laboratory, publishers (1988)). Production of MSP serum antibodies is also described in Chang et al., *Infection and Immunity* (1996) **64**:253-261 and Chang et al., (1992) *Proc Natl. Acad. Sci. USA* **86**:6343-6347.

The results of this analysis indicate that the modified MSP-1_{42} nucleic acid of the invention is expressed at a very high level compared to that of the natural protein which was not expressed at all. These results represent the first experimental evidence that reducing the AT % in a gene leads to expression of the MSP gene in heterologous systems and also the first evidence that removal of AUUUA mRNA instability motifs from the MSP coding region leads to the expression of MSP protein in COS cells.

Thus, the data presented here suggest that certain heterologous proteins that may be difficult to express in cell culture or transgenic systems because of high AT content and/or the presence of instability motifs, and or the usage of rare codons which are unrecognizable to the cell system of choice may be reengineered to enable expression in any given system with the aid of codon usage tables for that system. The present invention represents the first time that a DNA sequence has been modified with the goal of removing suspected sequences responsible for degradation resulting in low RNA levels or no RNA at all. The results shown in the Fig. 5, Panel A Northern (i.e. no RNA with native gene and reasonable levels with a modified DNA sequence in accordance with the invention), likely explains the increase in protein production.

The following examples describe the expression of MSP1-42 as a native non-fusion (and non-glycosylated) protein in the milk of transgenic mice.

Construction of MSP Transgene

To fuse MSP1-42 to the 15 amino acid ß-casein signal peptide, a pair of oligos, MSP203 and MSP204 (MSP203: ggccgctcgacgccaccatgaaggtcctcataattgcc tgtctggtggctctggccattgcagccgtcactccctccgtcat. MSP204: cgatgacggagggagtgacggctg caatggccagagccaccagacaggcaattatgaggaccttcatggtggcgtcgagc), which encode the 15 amino acid -

casein signal and the first 5 amino acid of the MSP1-42 ending at the Cla I site, was ligated with a Cla I-Xho I fragment of BC620 (Fig. 8) which encodes the rest of the MSP1-42 gene, into the Xho I site of the expression vector pCDNA3. A Xho I fragment of this plasmid (GTC669) was then cloned into the Xho I site of milk specific expression vector BC350 to generate B670 (Fig.9)

Expression of MSP1-42 in the milk of transgenic mice

A Sal I-Not I fragment was prepared from plasmid BC670 and microinjected into the mouse embryo to generate transgenic mice. Transgenic mice was identified by extracting mouse DNA from tail biopsy followed by PCR analysis using oligos GTC17 and MSP101 (sequences of oligos: GTC17, GATTGACAAGTAATACGCTGTTTCCTC, Oligo MSP101, GGATTCAATAGATACGG). Milk from the female founder transgenic mice was collected at day 7 and day 9 of lactation, and subjected to western analysis to determine the expression level of MSP-1-42 using an polyclonal anti-MSP antibody and monoclonal anti MSP antibody 5.2 (Dr. David Kaslow, NIH). Results indicated that the level of MSP-1-42 expression in the milk of transgenic mice was at 1-2 mg/ml (Fig. 10).

Construction of MSP1-42 glycosylation sites minus mutants

Our analysis of the milk produced MSP revealed that the transgenic MSP protein was N-glycosylated. To eliminate the N-glycosylation sites in the MSP1-42 gene, Asn. (N) at positions 181 and 262 were substituted with Gln.(Q). The substitutions were introduced by designing DNA oligos that anneal to the corresponding region of MSP1 and carry the AAC to CAG mutations. These oligos were then used as PCR primers to produce DNA fragments that encode the N to Q substitutions.

To introduce N262-Q mutation, a pair of oligos, MSPGYLYCO-3 (CAGGGAATGCTGCAGATCAGC) AND MSP42-2 (AATTCTCGAGTTAGTG GTGGTGGTGGTGGTGATCGCAGAAAATACCATG, FIG. 11), were used to PCR amplify plasmid GTC627, which contains the synthetic MSP1-42 gene. The PCR product was cloned into pCR2.1 vector (Invitrogen). This generated plasmid GTC716.

To introduce N181-Q mutation, oligos MSPGLYCO-1 (CTCCTTGTTCAGG AACTTGTAGGG) and MSPGLCO-2 (GTCCTGCAGTACACATATGAG, Fig 4) were used to

amplify plasmid GTC 627. The PCR product was cloned into pCR2.1. This generated plasmid GTC700.

The MSP double glycosylation mutant was constructed by the following three steps: first, a Xho I-Bsm I fragment of BC670 and the Bsm 1-Xho I fragment of GTC716 is ligated into the Xho I site of vector pCR2.1. This resulted a plasmid that contain the MSP-1-42 gene with N262-Q mutation. EcoN I-Nde I fragment of this plasmid was then replaced by the EcoN I-Nde I fragment from plasmid GTC716 to introduce the second mutation, N181-Q. A Xho I fragment of this plasmid was finally cloned into BC350 to generate BC718 (Fig. 12).

Expression of nonglycosylated MSP1in transgenic animals

BC718 has the following characteristics: it carries the MSP1-42 gene under the control of the ß-casein promoter so it can be expressed in the mammary gland of the transgenic animal during lactation. Further, it encodes a 15 amino acid ß-casein leader sequence fused directly to MSP1-42, so that the MSP1-42, without any additional amino acid at its N-terminal, can be secreted into the milk. Finally, because the N-Q substitutions, the MSP produced in the milk of the transgenic animal by this construct will not be N-glycosylated. Taken together, the transgenic MSP produced in the milk by BC718 is the same as the parasitic MSP.

A SalI/XhoI fragment was prepared from plasmid BC718 and microinjected into mouse embryos to generate transgenic mice. Transgenic animals were identified as described previously. Milk from female founders was collected and analyzed by Western blotting with antibody 5.2. The results, shown in Figure 13, indicate expression of nonglycosylated MSP1 at a concentration of 0.5 to 1 mg/ml.

Equivalents

Those skilled in the art will recognize, or be able to ascertain, using no more than routine experimentation, numerous equivalents are considered to be within the scope of this invention, and are covered by the following claims.

What is claimed is:

1. A modified known nucleic acid of a parasite which is capable of being expressed in a mammalian cell wherein the modification comprises a reduction of AT content of the gene by replacing one or more AT containing codons in the gene with a preferred codon encoding the same amino acid as the replaced codon.

2. A modified known nucleic acid of a parasite protein which is capable of being expressed in a mammalian cell wherein at least one mRNA instability motifs present in the gene coding sequence is eliminated by replacing said mRNA instability motif with a preferred codon encoding the same amino acid as the replaced codon.

3. The modified nucleic acid of claim 1 or 2 wherein at least one or more codons of the known gene is replaced by a preferred milk protein specific codon encoding the same amino acid as the replaced codon.

4. A modified known nucleic acid of a parasite which is capable of being expressed in a mammalian cell, wherein the overall AT content of the known gene encoding is lowered by replacement with a milk protein specific codon, and wherein at least one mRNA instability motif present in the gene is eliminated by replacement with a milk protein specific codon and at least one codon of the natural gene is replaced by a preferred milk protein specific codon.

5. The modified nucleic acid of claim 4 wherein said modified nucleic acid is capable of expressing said protein at a level which is at least 100% of that expressed by said natural gene in an *in vitro* or *in vivo* mammalian cell system.

6. A method for preparing a modified known nucleic acid of a parasite for expression in a mammalian cell comprising lowering the AT content of the natural gene by replacing one or more AT containing codons of the natural gene with a preferred mammary specific codon encoding the same amino acid as the replaced codon.

7. A method for preparing a modified known nucleic acid of a parasite protein for expression in a mammalian cell comprising eliminating at least one mRNA instability motif present in the gene coding sequence by replacing one or more mRNA instability motif in the gene with a mammary specific codon encoding the same amino acid as the replaced codon.

8. The method of claim 5 or 6 further comprising replacing one or more codons in the natural gene encoding said protein with a preferred mammary specific codon encoding the same amino acid as the replaced codon.

9. A modified nucleic acid sequence prepared by the method according to claim 5 or 6.

10. A method for preparing a modified known nucleic acid of a parasite for expression in a mammalian cell comprising the steps of:

a) eliminating at least one mRNA instability motif present in the natural gene encoding said protein by replacing one or more mRNA instability motifs in the gene with a preferred milk protein specific codon encoding the same amino acid as the replaced codon;

b) lowering the AT rich content of the natural gene encoding said protein by replacing one or more AT containing codons of the gene with a milk protein specific

codon encoding the same amino acid as the replaced codon; and

c) replacing one or more codons in the natural gene encoding said protein with a preferred mammary specific codon encoding the same amino acid as the replaced codon.

11. A modified nucleic acid prepared by the method according to claim 10.

12. A modified nucleic acid of claim 1 wherein said parasite is malaria and said nucleic acid is a fragment of SEQ ID NO 1 or SEQ ID NO 9 or a sequence specifically homologous thereto.

13. 12. A modified nucleic acid of claim 1 wherein said parasite is malaria and said nucleic acid is or SEQ ID NO 9 or a fragment thereof or a sequence specifically homologous thereto.

14. A modified nucleic acid that is a fragment of SEQ ID NO 1 or a sequence specifically homologous thereto capable of being expressed in a cell system wherein the AT content of the natural gene is lowered by replacement of one or more codons with codons recognizable by said cell culture system coding for the same amino acid as the replaced codon but which effectively lower the overall AT content of the natural gene.

15. A modified nucleic acid that is a fragment of SEQ ID NO 1 or a sequence specifically homologous thereto, capable of being expressed in a cell system wherein at least one mRNA instability motif present in the natural gene coding sequence is eliminated by replacing one or more codons comprising said instability motif with a codon recognizable by said cell system which effectively eliminates said instability motif and encodes the same amino acid as the replaced codon.

16. The modified nucleic acid of claims 14 or 15 wherein at least one or all codons of the natural gene are replaced with preferred codons of said cell system.

17. A vector comprising the modified nucleic acid of claim 12.

18. A host cell transfected or transformed with a vector of claim 17.

19. A transgenic expression construct comprising the modified nucleic acid of claim 12.

20. A transgenic non-human animal whose germline comprises the modified nucleic acid of claim 12.

21. A transgenic expression vector for the production of a transgenic animal comprising a promoter, operatively associated with the modified nucleic acid of claim 12, wherein said promoter directs mammary gland expression of the protein encoded by said modified nucleic acid into the animal's milk.

22. A modified known nucleic acid of a bacterium, virus, or parasite which is capable of being expressed in a cell system wherein the AT content of the gene is lowered by replacement of one or more codons with codons recognizable by said cell system coding for the same amino acid as the replaced codon, but which effectively lower the overall AT rich content of the natural gene.

23. A modified nucleic acid of a bacterium, virus, or parasite which is capable of being expressed in a cell system wherein at least one mRNA instability motifs present in the gene coding sequence is eliminated by replacing one or more codons comprising said instability motif with a codon recognizable by said cell system which

effectively eliminates said instability motif and encodes the same amino acid as the replaced codon.

24. A modified nucleic acid of claims 22 or 23, wherein at least one or all codons of the natural gene are replaced with preferred codons of said cell system.

25. A DNA vaccine comprising a modified nucleic acid according to claim 24.

26. A DNA vaccine comprising a vector according to claim 17.

Abstract of the disclosure

The invention provides modified recombinant nucleic acid sequences (preferably DNA) and methods for increasing the mRNA levels and protein expression of proteins which are known to be, or are likely to be, difficult to express in cell culture systems, mammalian cell culture systems, or in transgenic animals. The preferred "difficult" protein candidates for expression using the recombinant techniques of the invention are those proteins derived from heterologous cells preferably those of lower organisms such as parasites, bacteria, and virus, having DNA coding sequences comprising high overall AT content or AT rich regions and/or mRNA instability motifs and/or rare codons relative to the recombinant expression system to be used.

FIG. 1


1 GCAGTAACTCCTTCCGTAATT .AACATACTTTCTAAAATTGAAAATGAATA
1 AlaValThrProSerValIleAspAsnIleLeuSerLysIleGluAsnGluTy
EcoNI (73)
56 AGGTTTTATATTTAAAACCTTTAGCAGGTGTTTATAGAAGTTTAAAAAAACAATT
19 IuValLeuTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnLe
111 AGAAAATAACGTTATGACATTTAATGTTAATGTTAAGGATATTTTAAATTCACGA
37 uGluAsnAsnValMetThrPheAsnValAsnValLysAspIleLeuAsnSerArg
166 TTTAATAAACGTGAAAATTTCAAAAATGTTTTAGAATCAGATTTAATTCCATATA
56 PheAsnLysArgGluAsnPheLysAsnValLeuGluSerAspLeuIleProTyrL
221 AAGATTTAACATCAAGTAATTATGTTGTCAAAGATCCATATAAATTTCTTAATAA
74 ysAspLeuThrSerSerAsnTyrValValLysAspProTyrLysPheLeuAsnLy
276 AGAAAAAAGAGATAAATTCTTAAGCAGTTATAATTATATTAAGGATTCAATAGAT
92 sGluLysArgAspLysPheLeuSerSerTyrAsnTyrIleLysAspSerIleAsp
331 ACGGATATAAATTTTGCAAATGATGTTCTTGGATATTATAAAATATTATCCGAAA
111 ThrAspIleAsnPheAlaAsnAspValLeuGlyTyrTyrLysIleLeuSerGluL
386 AATATAAATCAGATTTAGATTCAATTAAAAAATATATCAACGACAAACAAGGTGA
129 ysTyrLysSerAspLeuAspSerIleLysLysTyrIleAsnAspLysGlnGlyGl
441 AAATGAGAAATACCTTCCCTTTTTAAACAATATTGAGACCTTATATAAAACAGTT
147 uAsnGluLysTyrLeuProPheLeuAsnAsnIleGluThrLeuTyrLysThrVal
496 AATGATAAAATTGATTTATTTGTAATTCATTTAGAAGCAAAAGTTCTAAATTATA
166 AsnAspLysIleAspLeuPheValIleHisLeuGluAlaLysValLeuAsnTyrT
551 CATATGAGAAATCAAACGTAGAAGTTAAAATAAAAGAACTTAATTACTTAAAAAC
184 hrTyrGluLysSerAsnValGluValLysIleLysGluLeuAsnTyrLeuLysTh
606 AATTCAAGACAAATTGGCAGATTTTAAAAAAAATAACAATTTCGTTGGAATTGCT
202 rIleGlnAspLysLeuAlaAspPheLysLysAsnAsnAsnPheValGlyIleAla
661 GATTTATCAACAGATTATAACCATAATAACTTATTGACAAAGTTCCTTAGTACAG
221 AspLeuSerThrAspTyrAsnHisAsnAsnLeuLeuThrLysPheLeuSerThrG
716 GTATGGTTTTTTGAAAATCTTGCTAAAACCGTTTTATCTAATTTACTTGATGGAAA
239 lyMetValPheGluAsnLeuAlaLysThrValLeuSerAsnLeuLeuAspGlyAs
771 CTTGCAAGGTATGTTAAACATTTCACAACACCAATGCGTAAAAAAACAATGTCCA
257 nLeuGlnGlyMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysPro
826 CAAAATTCTGGATGTTTCAGACATTTAGATGAAAGAGAAGAATGTAAATGTTTAT
276 GlnAsnSerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCysLeuL
881 TAAATTACAAACAAGAAGGTGATAAATGTGTTGAAAATCCAAATCCTACTTGTAA
294 euAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCysAs
936 CGAAAATAATGGTGGATGTGATGCAGATGCCAAATGTACCGAAGAAGATTCAGGT
312 nGluAsnAsnGlyGlyCysAspAlaAspAlaLysCysThrGluGluAspSerGly
991 AGCAACGGAAAGAAAATCACATGTGAATGTACTAAACCTGATTCTTATCCACTTT
331 SerAsnGlyLysLysIleThrCysGluCysThrLysProAspSerTyrProLeuP
PstI (1059)
1046 TCGATGGTATTTTCTGCAGTCACCACCACCACCACCACTAACT
349 heAspGlyIlePheCysSerHisHisHisHisHisHis•••

F16.2

FIG. 3A

Codon	AA	goat b-casein	goat K-casein	MSP wt	Edited MSP	mouse b-casein	mouse a-casein	mouse g-casein	mouse e-casein
TTT	Phe	5	4	8	0	4	8	3	4
TTC	Phe	4	3	7	15	4	6	7	1
TTA	Leu	0	2	25	0	0	0	0	0
TTG	Leu	0	2	3	0	0	0	0	1
TCT	Ser	5	1	4	1	13	5	7	5
TCC	Ser	2	2	2	3	6	14	8	2
TCA	Ser	1	4	10	1	1	3	2	0
TCG	Ser	0	1	0	0	0	0	0	0
TAT	Tyr	2	7	17	2	1	3	2	1
TAC	Tyr	1	2	3	18	2	6	6	7
TAA	***	1	2	0	0	1	0	1	0
TAG	***	0	0	0	0	0	0	0	0
TGT	Cys	1	1	10	12	0	0	1	0
TGC	Cys	0	2	2	0	2	2	2	1
TGA	***	0	0	0	0	0	1	0	1
TGG	Trp	1	1	0	0	0	2	2	2
CTT	Leu	9	1	9	0	16	9	3	3
CTC	Leu	5	2	0	0	7	8	0	1
CTA	Leu	1	2	1	0	1	2	1	0
CTG	Leu	11	5	0	38	10	17	4	1
CCT	Pro	17	6	4	2	8	6	3	0
CCC	Pro	12	0	1	6	8	6	6	4
CCA	Pro	3	13	5	1	5	6	2	2
CCG	Pro	1	1	0	1	0	0	0	1
CAT	His	0	1	3	0	2	6	2	1
CAC	His	5	3	1	4	4	0	3	0
CAA	Gln	5	9	9	0	9	21	9	7
CAG	Gln	16	6	0	9	21	32	12	8
CGT	Arg	0	1	1	0	0	0	0	0
CGC	Arg	0	0	0	0	1	0	0	0
CGA	Arg	0	0	1	0	0	0	0	1
CGG	Arg	1	0	0	3	0	0	0	0
ATT	Ile	4	5	13	0	3	4	3	4
ATC	Ile	6	3	2	20	7	5	8	5
ATA	Ile	1	3	5	0	1	0	2	0
ATG	Met	7	3	3	3	4	12	2	13
ACT	Thr	7	6	3	2	6	5	1	4
ACC	Thr	2	7	3	13	4	4	4	4
ACA	Thr	2	4	9	1	1	1	2	0
ACG	Thr	0	0	1	0	0	0	2	0
AAT	Asn	2	6	29	3	4	6	3	1
AAC	Asn	2	3	12	38	4	9	4	6
AAA	Lys	7	6	38	0	6	7	3	5
AAG	Lys	6	4	4	42	3	6	13	7
AGT	Ser	2	6	5	2	3	6	6	5
AGC	Ser	5	0	2	16	2	6	6	3
AGA	Arg	2	2	4	3	1	8	1	1
AGG	Arg	0	2	0	0	0	0	0	1
GTT	Val	5	6	15	0	7	4	2	3
GTC	Val	8	2	1	11	7	3	3	0
GTA	Val	2	2	5	0	2	4	1	3
GTG	Val	8	4	0	10	6	3	5	3
GCT	Ala	1	3	2	0	8	17	4	2
GCC	Ala	4	7	1	8	6	3	3	3
GCA	Ala	3	7	6	1	4	13	1	1
GCG	Ala	0	1	0	0	0	0	0	0
GAT	Asp	4	5	25	27	3	6	4	2
GAC	Asp	0	2	2	0	1	2	1	3
GAA	Glu	10	6	21	3	6	12	9	6
GAG	Glu	9	5	4	22	5	5	5	5
GGT	Gly	2	1	8	4	0	0	0	0
GGC	Gly	0	0	0	0	0	0	0	0
GGA	Gly	2	1	6	3	1	0	1	0
GGG	Gly	1	0	0	7	1	0	0	0

FIG. 3B

Codon	AA	MSP wt	Edited MSP	MSP wt	Edited MSP	E.coli	Human
TTT	Phe	8	0	0.53	0	0.5	0.35
TTC	Phe	7	15	0.47	1	0.5	0.65
TTA	Leu	25	0	0.66	0	0.11	0.05
TTG	Leu	3	0	0.08	0	0.11	0.09
TCT	Ser	4	1	0.17	0.04	0.27	0.17
TCC	Ser	2	3	0.09	0.13	0.21	0.26
TCA	Ser	10	1	0.43	0.04	0.13	0.11
TCG	Ser	0	0	0	0	0.14	0.07
TAT	Tyr	17	2	0.85	0.1	0.54	0.47
TAC	Tyr	3	18	0.15	0.9	0.46	0.53
TAA	***	0	0				
TAG	***	0	0				
TGT	Cys	10	12	0.83	1	0.45	0.3
TGC	Cys	2	0	0.17	0	0.55	0.7
TGA	***	0	0				
TGG	Trp	0	0	0	0	1	1
CTT	Leu	9	0	0.24	0	0.12	0.11
CTC	Leu	0	0	0	0	0.12	0.22
CTA	Leu	1	0	0.02	0	0.03	0.07
CTG	Leu	0	38	0	1	0.72	0.46
CCT	Pro	4	2	0.4	0.2	0.14	0.24
CCC	Pro	1	6	0.1	0.6	0.11	0.41
CCA	Pro	5	1	0.5	0.1	0.2	0.24
CCG	Pro	0	1	0	0.1	0.54	0.11
CAT	His	3	0	0.75	0	0.64	0.42
CAC	His	1	4	0.25	1	0.36	0.58
CAA	Gln	9	0	1	0	0.31	0.26
CAG	Gln	0	9	0	1	0.69	0.74
CGT	Arg	1	0	0.17	0	0.46	0.09
CGC	Arg	0	0	0	0	0.32	0.19
CGA	Arg	1	0	0.17	0	0.05	0.1
CGG	Arg	0	3	0	0.5	0.06	0.15
ATT	Ile	13	0	0.65	0	0.39	0.23
ATC	Ile	2	20	0.1	1	0.52	0.64
ATA	Ile	5	0	0.25	0	0.08	0.13
ATG	Met	3	3	1	1	1	1
ACT	Thr	3	2	0.19	0.13	0.36	0.2
ACC	Thr	3	13	0.19	0.81	0.38	0.47
ACA	Thr	9	1	0.56	0.06	0.09	0.21
ACG	Thr	1	0	0.06	0	0.17	0.12
AAT	Asn	29	3	0.71	0.07	0.29	0.34
AAC	Asn	12	38	0.29	0.93	0.71	0.66
AAA	Lys	38	0	0.9	0	0.72	0.45
AAG	Lys	4	42	0.1	1	0.28	0.55
AGT	Ser	5	2	0.21	0.09	0.11	0.11
AGC	Ser	2	16	0.09	0.7	0.14	0.29
AGA	Arg	4	3	0.67	0.5	0.08	0.24
AGG	Arg	0	0	0	0	0.03	0.23
GTT	Val	15	0	0.71	0	0.37	0.13
GTC	Val	1	11	0.05	0.52	0.12	0.27
GTA	Val	5	0	0.24	0	0.28	0.09
GTG	Val	0	10	0	0.48	0.23	0.5
GCT	Ala	2	0	0.22	0	0.33	0.31
GCC	Ala	1	8	0 11	0.89	0.18	0.4
GCA	Ala	6	1	0.67	0.11	0.28	0.17
GCG	Ala	0	0	0	0	0 21	0.12
GAT	Asp	25	27	0.93	1	0.48	0.38
GAC	Asp	2	0	0.07	0	0.52	0 62
GAA	Glu	21	3	0.84	0 12	0 67	0.4
GAG	Glu	4	22	0.16	0.88	0.33	0.6
GGT	Gly	8	4	0.57	0.29	0.46	0.15
GGC	Gly	0	0	0	0	0.4	0.44
GGA	Gly	6	3	0.43	0 21	0.06	0.17
GGG	Gly	0	7	0	0.5	0.08	0.24

FIG. 4A

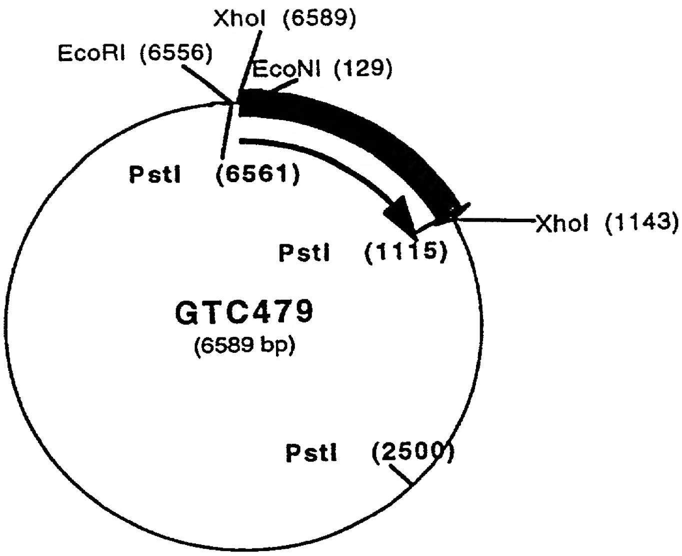

XhoI (6589)
EcoRI (6556)
EcoNI (129)
PstI (6561)
XhoI (1143)
PstI (1115)
GTC479
(6589 bp)
PstI (2500)

FIG. 4B

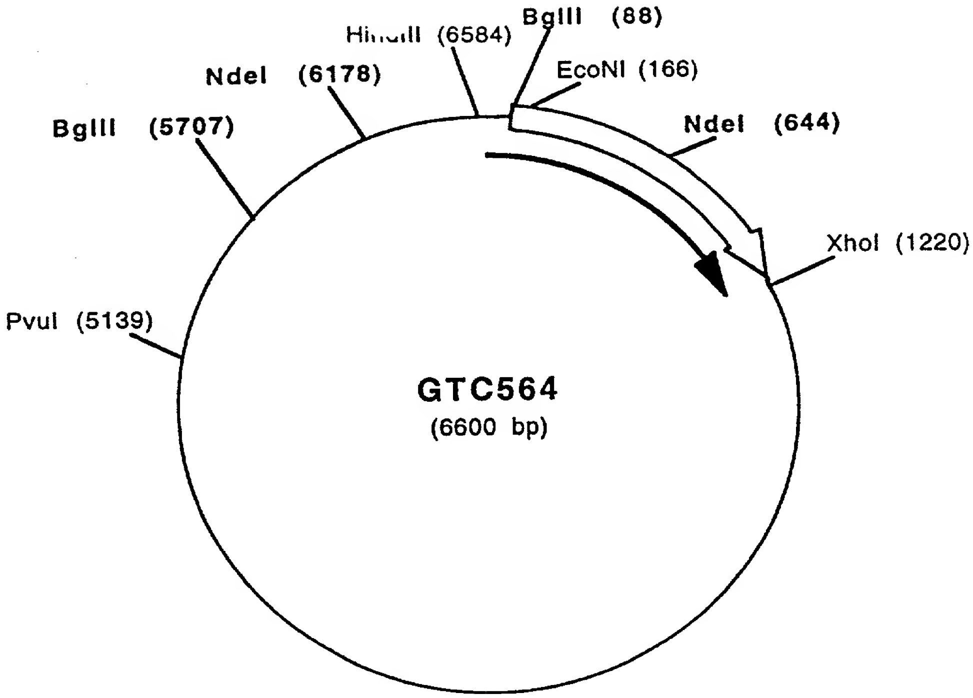


FIG. 4C

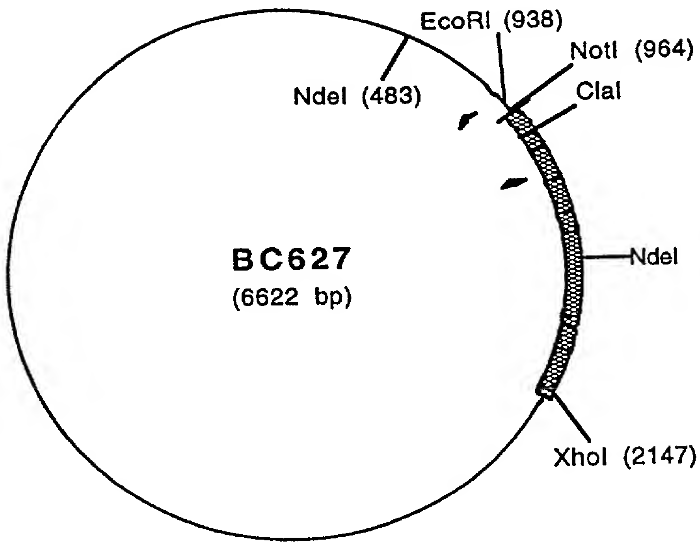
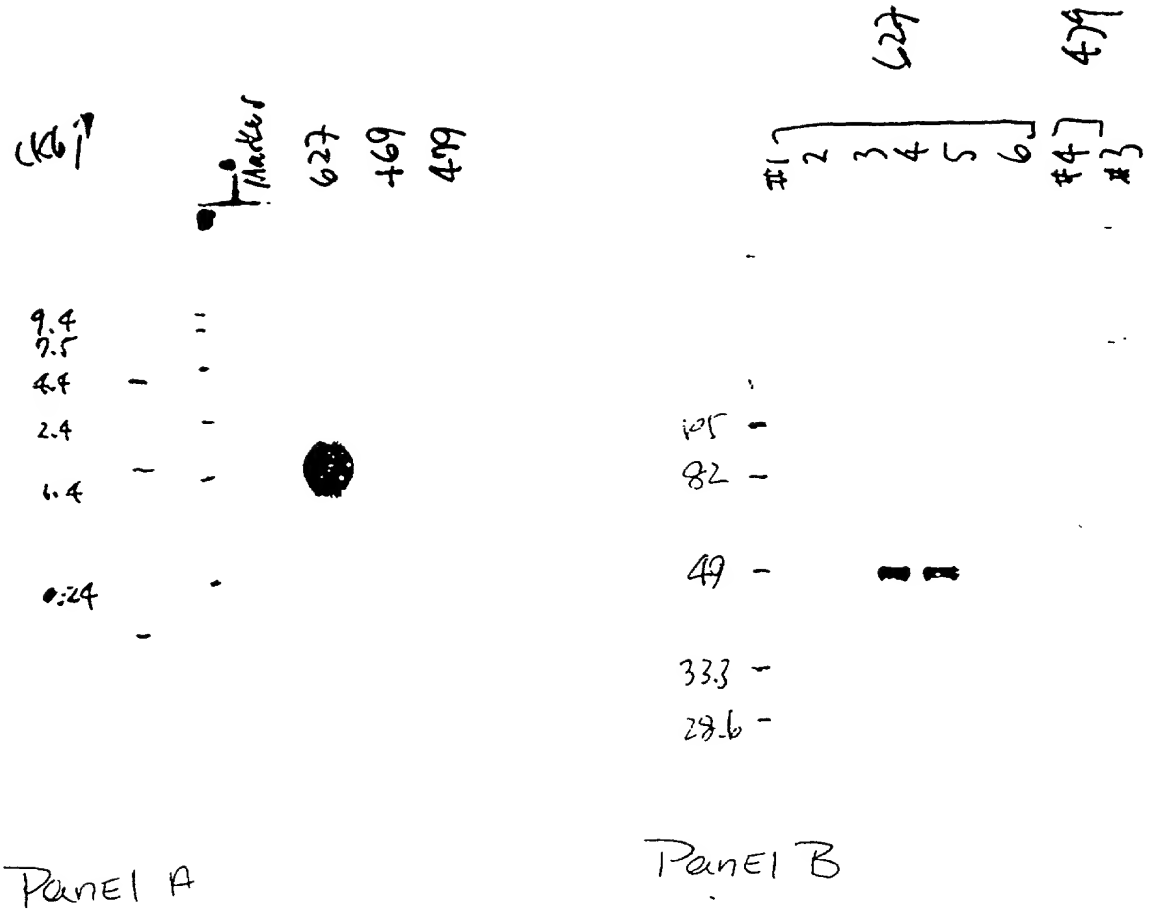


FIG. 5



Panel A

Panel B

FIG. 6

Oligos used:

OT1:
TCG ACG AGA GCC ATG AAG GTC CTC ATC CTT GCC TGT CTG GTG GCT CTG GCC ATT GCA AGA GAG CAG GAA GAA CTC AAT GTA GTC GGT A,

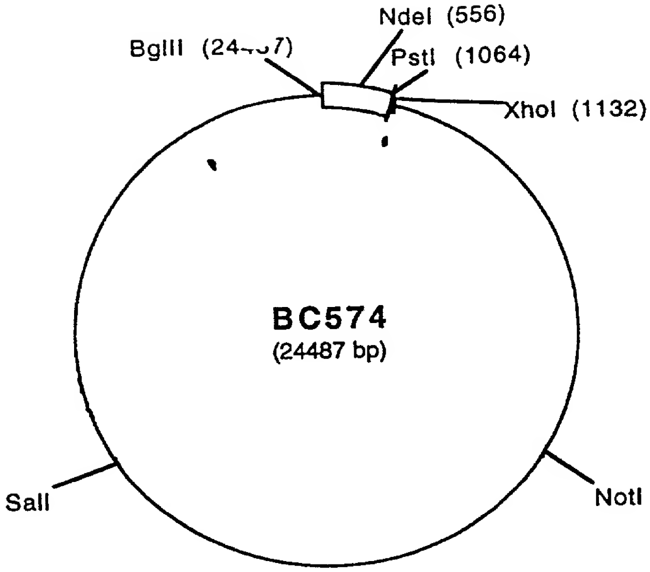
OT2:
GAT CTA CCG ACT ACA TTG AGT TCT TCC TGC TCT CTT GCA ATG GCC AGA GCC ACC AGA CAG GCA AGG ATG AGG ACC TTC ATG GCT CTC G,

MSP1:
AATAGATCTGCAGTAACTCCTTCCGTAATTG,

MSP2:
AATTCTCGAGTTAGTGGTGGTGGTGGTGGTGACTGCAGAAATACCATC

MSP8:
TAACTCGAGCGAACCATGAAGGTCCTCATCCTTGCCTGTCTGGTGGCTCTGGCCATTGCA

FIG. 7


NdeI (556)
BglII (24[illegible]7)
PstI (1064)
XhoI (1132)
BC574
(24487 bp)
SalI
NotI

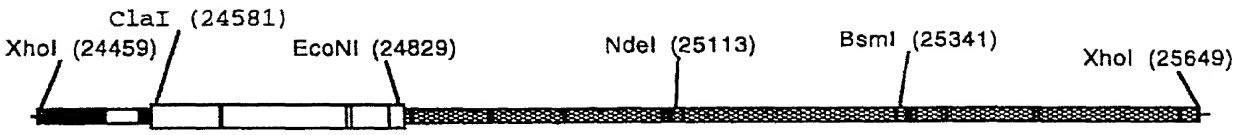


Diagram of BC620

FIGURE 8

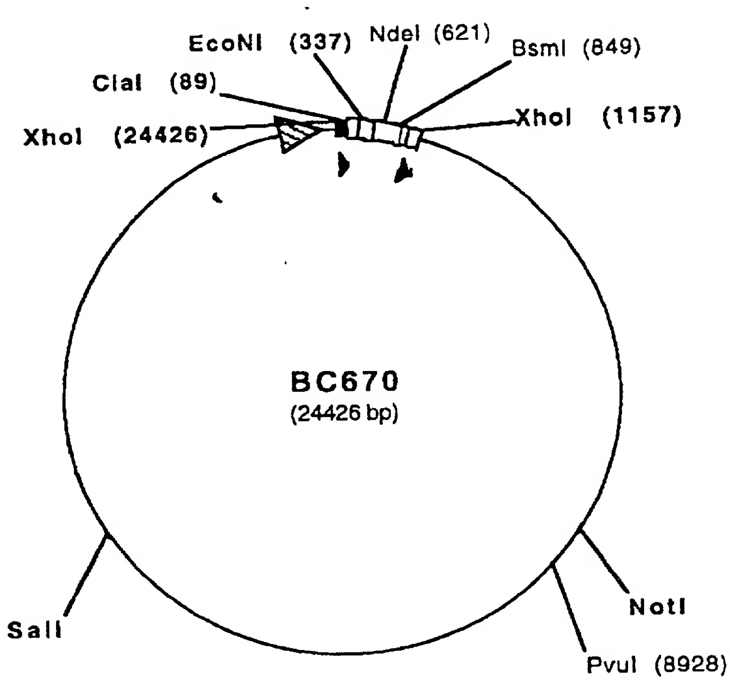
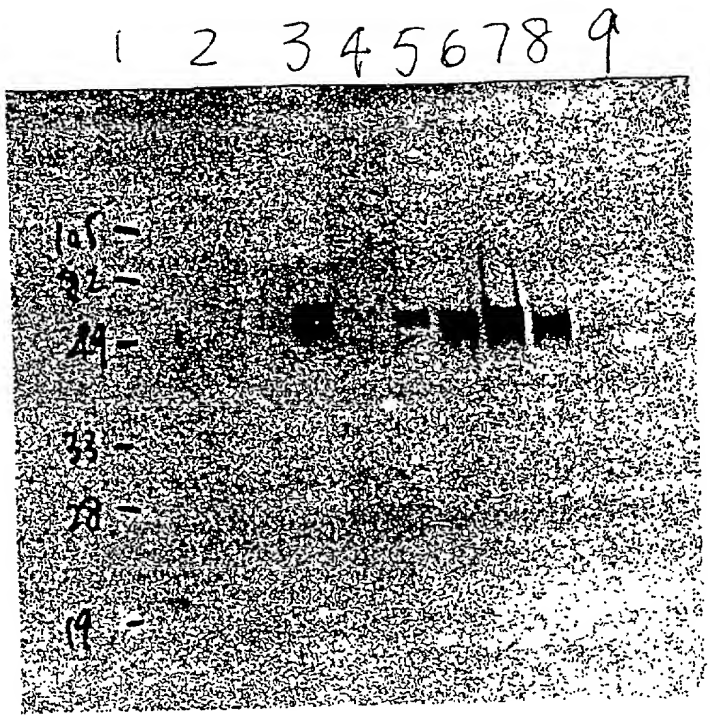


FIGURE 9



Western Analysis of MSP transgenic milk.
Lane 1, Molecular weight marker; lane 2, nontransgenic mice milk; lane 3, milk from BC628-146 transgenic mouse; lane 4-9, milk from BC670 transgenic mice. The blot was reacted with monoclonal antibody 5.2 against MSP.

FIGURE 10

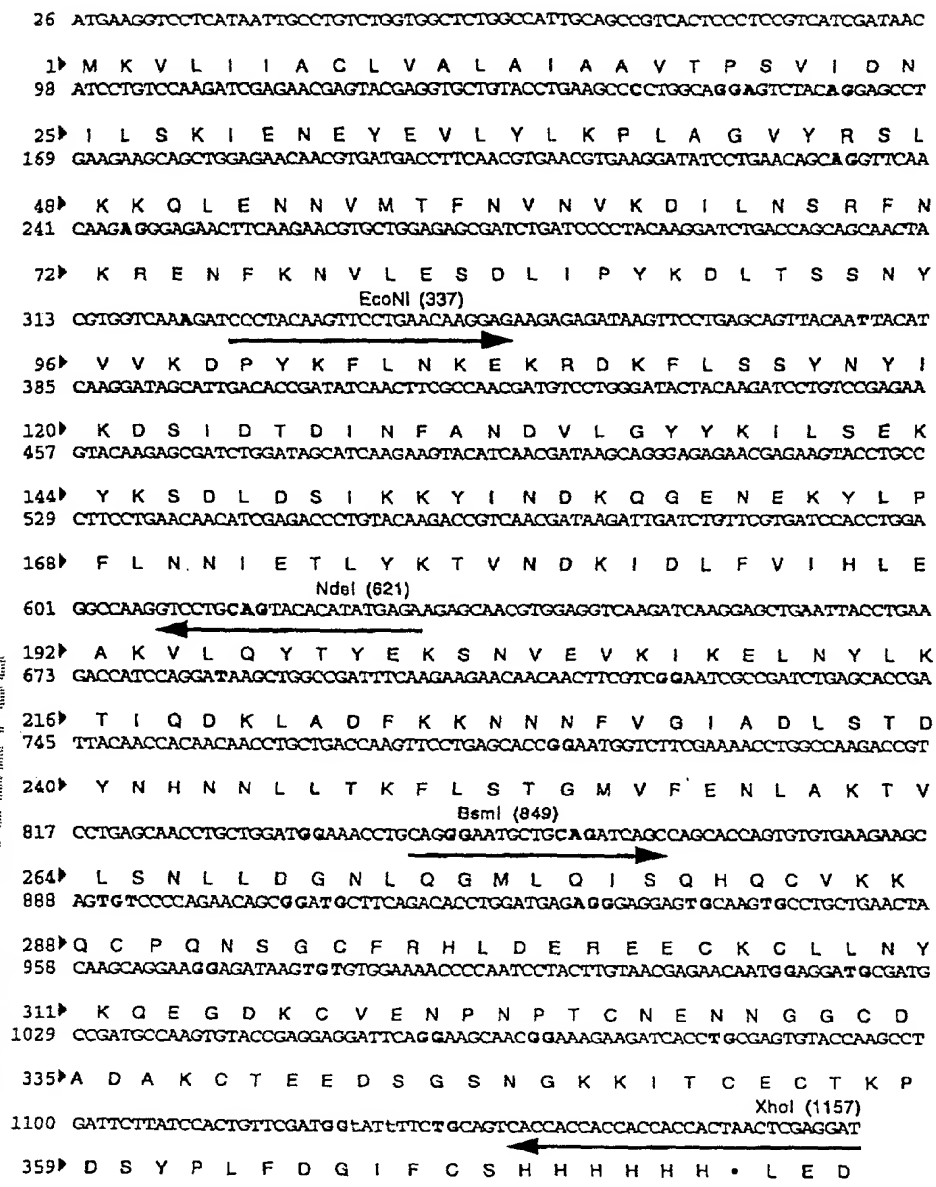

26 ATGAAGGTCCTCATAATTGCCTGTCTGGTGGCTCTGGCCATTGCAGCCGTCACTCCCTCCGTCATCGATAAC
1 M K V L I I A C L V A L A I A A V T P S V I D N
98 ATCCTGTCCAAGATCGAGAACGAGTACGAGGTGCTGTACCTGAAGCCCCTGGCAGGAGTCTACAGGAGCCT
25 I L S K I E N E Y E V L Y L K P L A G V Y R S L
169 GAAGAAGCAGCTGGAGAACAACGTGATGACCTTCAACGTGAACGTGAAGGATATCCTGAACAGCAGGTTCAA
48 K K Q L E N N V M T F N V N V K D I L N S R F N
241 CAAGAGGGAGAACTTCAAGAACGTGCTGGAGAGCGATCTGATCCCCTACAAGGATCTGACCAGCAGCAACTA
72 K R E N F K N V L E S D L I P Y K D L T S S N Y
EcoNI (337)
313 CGTGGTCAAAGATCCCTACAAGTTCCTGAACAAGGAGAAGAGAGATAAGTTCCTGAGCAGTTACAATTACAT
96 V V K D P Y K F L N K E K R D K F L S S Y N Y I
385 CAAGGATAGCATTGACACCGATATCAACTTCGCCAACGATGTCCTGGGATACTACAAGATCCTGTCCGAGAA
120 K D S I D T D I N F A N D V L G Y Y K I L S E K
457 GTACAAGAGCGATCTGGATAGCATCAAGAAGTACATCAACGATAAGCAGGGAGAGAACGAGAAGTACCTGCC
144 Y K S D L D S I K K Y I N D K Q G E N E K Y L P
529 CTTCCTGAACAACATCGAGACCCTGTACAAGACCGTCAACGATAAGATTGATCTGTTCGTGATCCACCTGGA
168 F L N N I E T L Y K T V N D K I D L F V I H L E
NdeI (621)
601 GGCCAAGGTCCTGCAGTACACATATGAGAAGAGCAACGTGGAGGTCAAGATCAAGGAGCTGAATTACCTGAA
192 A K V L Q Y T Y E K S N V E V K I K E L N Y L K
673 GACCATCCAGGATAAGCTGGCCGATTTCAAGAAGAACAACAACTTCGTCGGAATCGCCGATCTGAGCACCGA
216 T I Q D K L A D F K K N N N F V G I A D L S T D
745 TTACAACCACAACAACCTGCTGACCAAGTTCCTGAGCACCGGAATGGTCTTCGAAAACCTGGCCAAGACCGT
240 Y N H N N L L T K F L S T G M V F E N L A K T V
BsmI (849)
817 CCTGAGCAACCTGCTGGATGGAAACCTGCAGGGAATGCTGCAGATCAGCCAGCACCAGTGTGTGAAGAAGC
264 L S N L L D G N L Q G M L Q I S Q H Q C V K K
888 AGTGTCCCCAGAACAGCGGATGCTTCAGACACCTGGATGAGAGGGAGGAGTGCAAGTGCCTGCTGAACTA
288 Q C P Q N S G C F R H L D E R E E C K C L L N Y
958 CAAGCAGGAAGGAGATAAGTGTGTGGAAAACCCCAATCCTACTTGTAACGAGAACAATGGAGGATGCGATG
311 K Q E G D K C V E N P N P T C N E N N G G C D
1029 CCGATGCCAAGTGTACCGAGGAGGATTCAGGAAGCAACGGAAAGAAGATCACCTGCGAGTGTACCAAGCCT
335 A D A K C T E E D S G S N G K K I T C E C T K P
XhoI (1157)
1100 GATTCTTATCCACTGTTCGATGGTATTTCTGCAGTCACCACCACCACCACCACTAACTCGAGGAT
359 D S Y P L F D G I F C S H H H H H H • L E D

FIGURE 11

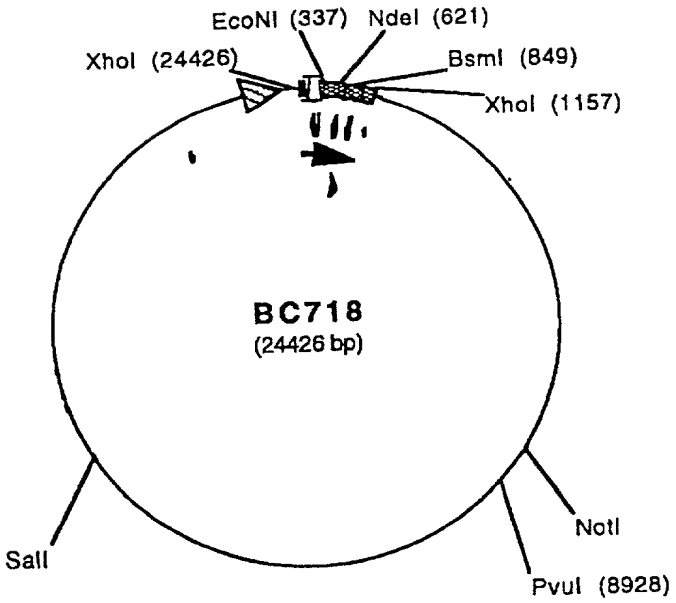
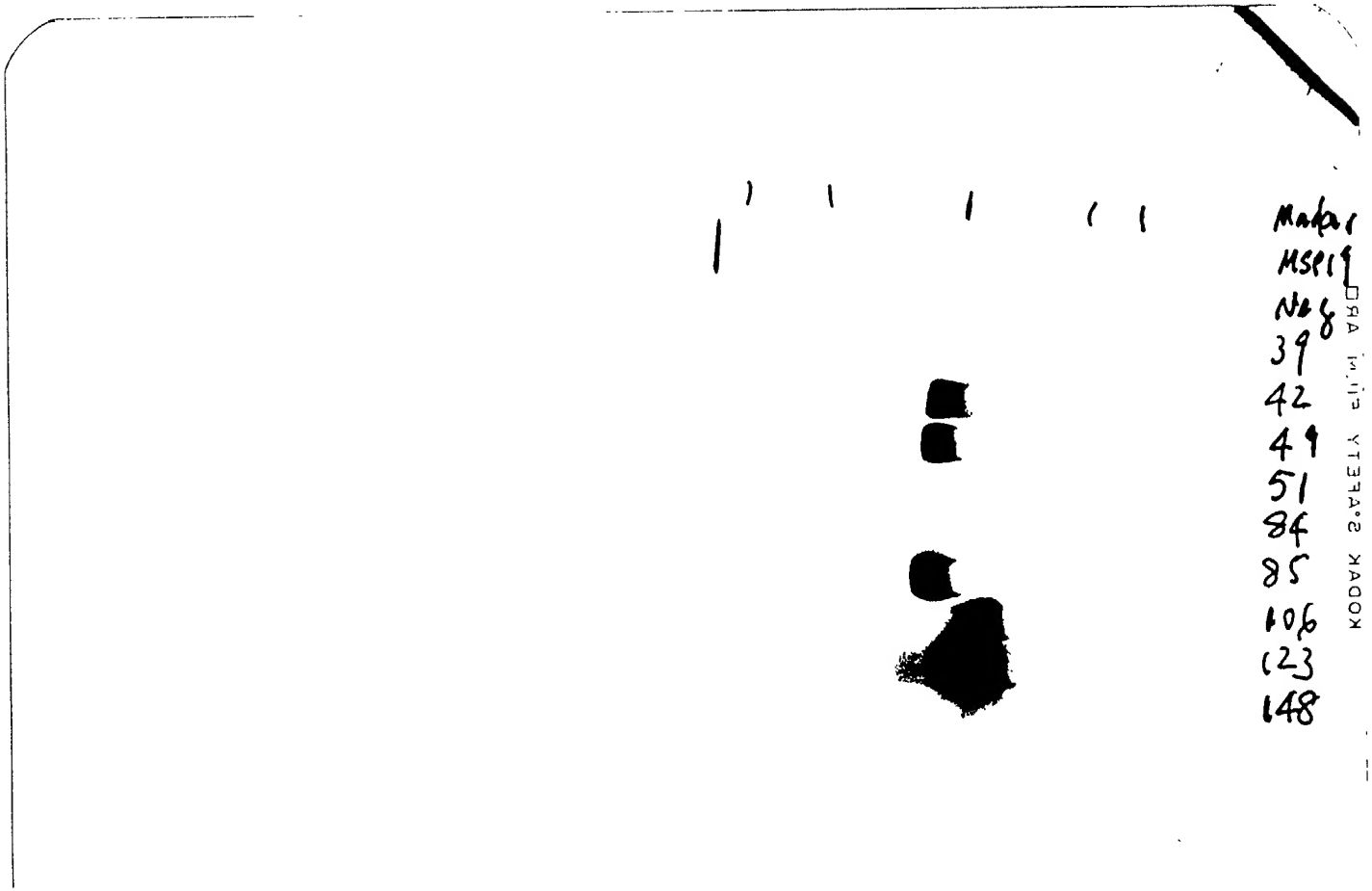


FIGURE 12

FIGURE 13



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
(Case No. 107.637.121-B)

Applicant or Patentee: Li How Chen and Harry Meade

Serial No.

Filed or Issued: HEREWITH

For: NOVEL MODIFIED NUCLEIC ACID SEQUENCES AND METHODS FOR INCREASING mRNA LEVELS AND PROTEIN EXPRESSION IN CELL SYSTEMS

VERIFIED STATEMENT (DECLARATION) CLAIMING SMALL ENTITY STATUS (37 C.F.R. § 1.9(f) AND § 1.27(c)) - SMALL BUSINESS CONCERN

I hereby declare that I am an official of the small business concern empowered to act on behalf of the concern identified below:

NAME OF SMALL BUSINESS CONCERN: Genzyme Transgenics Corporation
ADDRESS OF SMALL BUSINESS CONCERN: 5 Mountain Road, Framingham, MA 01701

I hereby declare that the above-identified small business concern qualifies as a small business concern as defined in 13 C.F.R. § 121.12 and reproduced in 37 C.F.R. § 1.9(d), for purposes of paying reduced fees to the United States Patent and Trademark Office, in that the number of employees of the business concern, including those of its affiliates, does not exceed 500 persons. For purposes of this statement (1) the number of employees of the business concern is the average over the previous fiscal year of the concern of the persons employed on a full-time, part-time or temporary basis during each of the pay periods of the fiscal year, and (2) concerns are affiliates of each other when either, directly or indirectly, one concern controls or has the power to control the other, or a third part or parties controls or has the power to control both.

I hereby declare that the rights under contract or law have been conveyed to and remain with the small business concern identified above with regard to the invention, entitled

NOVEL MODIFIED NUCLEIC ACID SEQUENCES AND METHODS FOR INCREASING mRNA LEVELS AND PROTEIN EXPRESSION IN CELL SYSTEMS

by inventors, Li How Chen and Harry Meade, described in

[X] the specification filed herewith

[ ] Application Serial No. _____, filed _____.

[ ] Patent No. _____, issued _____.

If the rights held by the above-identified small business concern are not exclusive, each individual, concern or organization having rights to the invention is listed below* and no rights to the invention are held by any person, other than the inventor, who would not qualify as an independent inventor under 37 C.F.R. § 1.9(c) if that person made the invention, or by any concern which would not qualify as a small business concern under 37 C.F.R. § 1.9(d), or a nonprofit organization under 37 C.F.R. § 1.9(e).

FULL NAME ______________________

ADDRESS ______________________

[ ] Individual [ ] Small Business Concern [ ] Nonprofit Organization

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 C.F.R. § 1.28(b))

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing therein, or any patent to which this verified statement is directed.

____

NAME OF PERSON SIGNING

____

TITLE IN ORGANIZATION

5 Mountain Road, Framingham, MA 01701

ADDRESS OF PERSON SIGNING

____

Signature

Date: ______________________

*Note: Separate verified statements are required from each named person, concern or organization having rights to the invention averring to their status as small entities. (37 C.F.R. § 1.27).